

JLIS.it
Italian Journal of Library and Information Science
Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione

Analisi e riflessioni sul VIAF, Virtual International Authority File

Giulia Manzotti

Premessa

Tra i numerosi progetti internazionali emersi nell'ultimo ventennio, il Virtual International Authority File (VIAF)[1] si configura senz'altro come l'esperimento più importante nato in seno alla cooperazione internazionale per quel che riguarda l'authority control.

A ottobre 2010, sono diciotto le agenzie catalografiche partecipanti al progetto che, con l'apporto degli authority record provenienti dai propri authority file, consentono la messa a punto di un servizio di ricerca di voci d'autorità sempre più esaustivo.[2]

Il VIAF, concepito come progetto nei primi anni Duemila, è un esperimento congiunto della Library of Congress (LC), della Deutsche Nationalbibliothek (DNB) e della Bibliothèque Nationale de

[1] http://www.viaf.org.

[2] National Library of Australia, National Library of the Czech Republic, Bibliotheca Alexandrina (Egitto), Bibliothèque Nationale de France, Deutsche Nationalbibliothek, Getty Research Institute, National Library of Israel, ICCU (Italia), Biblioteca Nacional de Portugal, Biblioteca Nacional de España, National Library of Sweden, Swiss National Library, Library of Congress/NACO, Vatican Library, NUKAT Center (Polonia), Library and Archives Canada, NII (Giappone), National Széchényi Library (Ungheria). È in fase di testing il caricamento di authority data della Rete delle biblioteche della Svizzera occidentale (RERO).

JLIS.it. Vol. 1, n. 2 (Dicembre/December 2010), p. 357–381.
DOI: 10.4403/jlis.it-4520



France (BNF) con Online Computer Library Center (OCLC); obiettivo precipuo del VIAF è il collegamento automatico, mediante lo sviluppo di algoritmi di schiacciamento (matching algorithm), degli authority record delle agenzie nazionali partecipanti, rendendoli poi disponibili in rete attraverso un authority service globale e gratuito. Attraverso la propria interfaccia, il VIAF permette la ricerca virtuale di voci d'autorità nella lingua, scrittura e alfabeto utilizzati dall'utente. Gli authority record attualmente disponibili[3] sono quelli relativi ai nomi personali, sebbene OCLC, responsabile della parte tecnico-informatica del progetto, abbia sviluppato degli algoritmi di schiacciamento per il collegamento automatico dei nomi geografici; tuttavia la percentuale del tasso d'errore nei collegamenti degli stessi risulta ancora troppo alta per essere accettata. Da parte dei gestori del VIAF è fortemente sentita, inoltre, l'esigenza di allargare il servizio anche ai nomi di ente, sia a carattere permanente che temporaneo.

L'implementazione di un servizio quale il VIAF, che permette una ricerca in più lingue e scritture di voci d'autorità nel rispetto della nazionalità dell'utente, è il risultato sul piano pratico di diversi dibattiti teorici che hanno caratterizzato gli ultimi venti anni del secolo scorso.

Retroterra storico

Il progetto VIAF nacque nei primi anni Duemila in un clima di forte sensibilizzazione verso una più specifica definizione di authority control, in particolar modo nei significati che questo poteva assumere quando applicato a livello internazionale. Già nel 1978, in occasione della 44ª Sessione dell'IFLA in Strbské Pleso (Cecoslovacchia), era stata concepita l'idea di sviluppare un authority

[3] Ottobre 2010.

system internazionale, per la cui implementazione venne creato un apposito gruppo di lavoro. Tra gli obiettivi perseguiti dal gruppo, si segnalavano la discussione e formulazione delle specifiche per un authority system internazionale, finalizzato a soddisfare le esigenze bibliografiche delle biblioteche, lo sviluppo di un formato UNIMARC per authorities[4] e metodi di scambio efficienti ed efficaci di authority data.

Il primo obiettivo, che può condensarsi nell'elaborazione di standard che stabiliscano forma, ordine e punteggiatura nelle intestazioni, fu concretizzato nel 1984 con la pubblicazione delle Guidelines for authority and reference entries (GARE). Le linee guida fornivano una definizione per la struttura delle intestazioni al livello più generale, poiché la determinazione della forma delle intestazioni continuava a essere prerogativa di ogni singola agenzia catalografica nazionale. Nel 2001, dopo essere state sottoposte a processo di revisione e ampliate, le Guidelines furono riedite con il nome di Guidelines for Authority and Reference Records (GARR). Le GARR estendevano la loro applicazione a un maggior numero di supporti documentari, e regolamentavano la visualizzazione dei dati d'autorità e dei rinvii per tutti i tipi di materiali.

Le *Guidelines* erano state costruite sul concetto di controllo bibliografico universale, ma la presa di coscienza della comunità degli studiosi sulla necessità di conservazione delle tradizioni culturali di ciascuna agenzia nazionale era sempre più forte. Tale necessità poteva essere soddisfatta grazie alle possibilità offerte dalle nuove tecnologie: consentendo il collegamento automatico degli authority record delle varie biblioteche, l'utilizzo di un'unica forma dell'intestazione per tutti i paesi non risultava più una via obbligata da seguire. Grazie a questo nuovo approccio, l'utente sarebbe stato

[4]Lo sviluppo del formato UNIMARC per i dati di autorità fu affidato, nel 1984, a uno Steering Group dell'IFLA.

messo nella condizione di effettuare ricerche o di visualizzare le intestazioni nella propria lingua, scrittura o alfabeto.

Su questo principio nacquero i primi progetti internazionali legati all'authority control, come ad esempio AUTHOR (1995), che mirava all'interoperabilità e allo scambio di authority record tra cinque biblioteche nazionali europee. Ciascuna agenzia continuava a essere responsabile dei propri authority file, a usare le proprie regole di catalogazione, i propri formati di codifica dei record, nonché la propria lingua: lo scambio di record avveniva utilizzando dei convertitori automatici. L'accesso in rete ai vari authority file era possibile sfruttando il lavoro di identificazione degli autori a livello locale.

Da questo progetto, conclusosi nel 1998, nacque la proposta di definire un formato minimo per i record di autorità. Parallelamente a questa ricerca scientifica ed editoriale, volta alla definizione di standard e alla realizzazione di progetti prototipali per il collegamento automatico di authority record tra le diverse biblioteche, il dibattito teorico sul significato di authority control si era fatto anch'esso molto vivace. Nuove tendenze, intese a ridefinire il concetto di authority control, erano emerse all'interno di numerosi gruppi di lavoro dell'IFLA già alla fine degli anni Novanta: il concetto di intestazione uniforme, presente nelle normative catalografiche, cui si era guardato fino a quel momento, appariva limitativo (se non sbagliato) quando applicato a livello internazionale.

L'utilizzo di un'unica intestazione, detta autorevole, valida per tutti i paesi e quindi utilizzata globalmente e universalmente, doveva lasciar spazio, perlomeno in ambito internazionale, all'utilizzo di strutture sindetiche che raggruppano, attraverso un meccanismo di collegamento, tutte le forme varianti della medesima entità (Guerrini e Sardo 35). Secondo questo modello, le forme varianti di un nome (siano esse linguistiche, grafiche, nell'ordine degli elementi)

possono convivere all'interno del medesimo record senza entrare in conflitto tra loro, poiché nessuna di esse è dichiarata come autorevole; in questo modo, viene altresì assicurato il rispetto delle diversità linguistiche e culturali dei singoli paesi. Quanto appena esposto è alla base di un autorevole studio condotto da Barbara Tillett su un'adeguata definizione di authority control: considerando il problema dal punto di vista dello scambio internazionale delle informazioni, la studiosa propose un modello strutturale denominato access point record, identificandolo come possibile materializzazione di un authority record internazionale (Tillett, «Access control: a model for descriptive, holding and control records»). Punto chiave dell'access point record è l'abbandono della nozione di autorità/autorevolezza cui è legato il concetto di authority control; l'access point record è in grado di mantenere al proprio interno tutte le forme preferite stabilite da ciascuna agenzia bibliografica, senza dichiararne una come autorevole. In questo modo, tutte le forme varianti sarebbero collegate a una sola registrazione di controllo. Il modello proposto da Tillett risponde a esigenze di ordine pragmatico, che riconoscono che un utente o una biblioteca vorrebbero poter cercare, o scegliere, la forma preferita per un nome secondo la propria lingua, scrittura o alfabeto. La necessità di scambio di authority data a livello internazionale e, a questo scopo, il parallelo bisogno di rispetto delle diversità linguistiche e culturali dei singoli paesi, hanno dato la spinta decisiva alla realizzazione del progetto VIAF.

Definizione del progetto e background tecnico

Sebbene il principio ispiratore del VIAF fosse stato già preannunciato in alcuni contributi di Barbara Tillett dei primi anni Duemi-

la («Authority Control on the Web»;«A Virtual International Authority File»), fu in occasione dell'IME ICC di Francoforte del 2003 che Barbara Tillett annunciava (Tillett, Gömpel e Oehlschläger) l'avvio all'implementazione del VIAF, comunicando come parecchi gruppi di lavoro all'interno dell'IFLA, sezione di catalogazione, avessero riconosciuto la potenzialità offerta da un "virtual international authority file" (Tillett, Lee e Cristán).

Durante il 72º Congresso IFLA di Seul del 2006, furono presentati per la prima volta gli sviluppi concreti del progetto (Bennett et al.), attraverso l'intervento di alcuni membri della LC, di OCLC e di DNB. Gli obiettivi perseguiti nella realizzazione della versione iniziale del VIAF furono essenzialmente cinque:

1. costruzione degli enhanced record (lett. "record migliorati"). Questa prima fase consisteva in un'identificazione manuale degli authority record più appropriati da includere negli enhanced record; i dati necessari sarebbero stati estratti dall'authority file tedesco, il Personennormdatei (PND), e dal Library of Congress Name Authority File (LCNAF);

2. sviluppo di algoritmi di schiacciamento (matching algorithm) e confronto tra gli enhanced record del PND e del LCNAF per creare la versione iniziale del VIAF;

3. utilizzo del protocollo della Open Archives Iniziative (OAI), con un server centrale che raccolga i metadati degli authority file delle agenzie nazionali o regionali. Grazie a tale sistema, il VIAF è stato costruito sugli authority file già formati di ciascuna agenzia bibliografica, sfruttando il lavoro fatto a livello locale. In altri termini, il VIAF è virtuale proprio perché l'authority record completo rimane nell'authority file di appartenenza, mentre i dati raccolti nel server centrale sono minimi. I record relativi alla stessa entità sono collegati al ser-

ver centrale e aggiornati ogniqualvolta potrebbero verificarsi dei cambiamenti nei file nazionali;

4. mantenimento del VIAF database. Per permettere l'aggiornamento e il mantenimento del sistema, venivano richieste a tutte le agenzie partecipanti, nonché maggiormente alla LC e a DDB, periodiche aggiunte e variazioni ai record; ogni biblioteca doveva continuare le proprie attività di gestione giornaliera. Il sistema di richiesta di aggiornamento sarebbe stato costruito sui protocolli usati dall'OAI server;

5. accesso ai VIAF record attraverso un'interfaccia utente disponibile nell'open Web. Il database e l'interfaccia avrebbero supportato Unicode, nonché ricerche in più lingue e scritture, elementi di vera innovazione introdotti dal VIAF. Per sfruttare le capacità del cosiddetto "Web semantico" (Berners-Lee, Hendler e Lassila), sarebbe stato reso possibile, partendo da una forma del nome accettata dalla LC, cliccare su un link HTML ottenendone la versione corrispondente della DNB.

Ai suoi esordi, il VIAF si configurava come un vero e proprio dizionario Tedesco-Inglese e Inglese-Tedesco per i nomi personali. Qualora un utente americano avesse cercato l'autore J. P. De Valk (forma accettata dalla LC), a fronte della sua query avrebbe ricevuto automaticamente la "traduzione" del nome cercato, ovvero Johannes P. de Valk (forma accettata dalla DNB). Questo esempio è utile per affrontare una tematica non scontata e che crea numerosi problemi negli authority file nazionali: spesso le varie agenzie nazionali, come nel caso sopracitato, scelgono forme del nome differenti per la medesima persona o, viceversa, usano la stessa forma del nome per rappresentare autori diversi. È possibile che J. P. de Valk sia stato attribuito dalla DNB a un autore completamente diverso da quello della LC.

Per confermare ed evidenziare maggiormente alla comunità di studiosi la difficoltà di schiacciare authority record senza usare informazioni supplementari oltre al nome, la LC e la DNB estrassero dai rispettivi authority file un campione di nomi comuni. Le coppie di nomi vennero controllate manualmente per determinare se rappresentassero effettivamente la stessa persona: da questa revisione, le due agenzie trovarono che circa il 10% degli accoppiamenti di nomi personali si riferivano a persone diverse (Bennett et al. 5). Capiamo quindi che il tasso di errore nell'abbinamento, qualora venga usata solamente la forma stabilita del nome, è troppo alto per essere accettato.

La soluzione per ovviare al problema appena esposto fu individuata nelle informazioni addizionali legate al nome personale, risultando chiaramente necessarie per confermare o respingere potenziali schiacciamenti di nomi. I titoli dell'opera, ad esempio, se correttamente attribuiti a un autore a monte, ovvero negli authority file delle varie agenzie nazionali, forniscono un supporto quasi definitivo per lo schiacciamento di un nome. Tuttavia, quando il titolo di un'opera in uno dei due authority data si presenta tradotto in un'altra lingua, difficilmente l'elemento titolo del record può essere utilizzato come supporto allo schiacciamento automatico; in questo caso, può essere maggiormente utile la presenza, nel record, di un coautore.

Per la conferma di un match, l'approccio tenuto dai gestori del VIAF implicava che:

1. i nomi dovessero essere compatibili;

2. dovesse esserci almeno un'informazione supplementare sufficiente di conferma.

Durante il test di compatibilità, venivano prese in considerazione sia la forma autorizzata che le forme varianti del nome, qualora

fossero presenti; una volta che i nomi venivano dichiarati come compatibili, per confermare lo schiacciamento il sistema utilizzava le informazioni supplementari. I file bibliografici possono contenere tanti titoli diversi ma simili, come anche tanti nomi differenti ma simili. La strategia di base seguita dal VIAF era la seguente: se una coppia nome/titolo è simile in entrambi i file, allora il nome rappresenta la stessa persona. Le date presenti nei record erano trattate separatamente come una *positive correlation* (lett.: correlazione positiva). Quando le date differivano per più di un anno, i nomi erano considerati incompatibili, e quindi il match respinto. Erano giustificate, invece, le differenze di un solo anno. Durante lo sviluppo del VIAF, fu relativamente facile trovare piccole discrepanze in alcune datazioni, tuttavia l'informazione supplementare era sufficiente a confermare il match sebbene vi fossero queste leggere variazioni.

Per un maggiore approfondimento, è utile dire che i punti di schiacciamento venivano distinti in tre categorie: forte, moderato e debole. I punti "forti" di schiacciamento per due nomi che sono già stati determinati come compatibili sono: i titoli, gli ISBN, le date di nascita e di morte, i coautori. La sola data di nascita non era considerata sufficiente per disambiguare i nomi, per cui veniva propriamente considerata come un punto di schiacciamento "moderato". Altri punti di abbinamento moderati sono gli indicatori dell'ambiente di lavoro delle persone o il ruolo svolto dalla persona (illustratore, compositore etc.). Punti di schiacciamento "deboli" includono la lingua, l'area di soggetto e il paese di pubblicazione. Per combinare i punti di schiacciamento, a ciascuno di essi viene attribuito un punteggio numerico. Per punti di schiacciamento composti da cifre come ad esempio un ISBN, il match o è esatto (e gli viene attribuito un punteggio di 1) o non è un match (punteggio 0). Per punti di schiacciamento costituiti da testo, come i titoli, poteva essere assegnato un punteggio in base a quanto fosse simile il testo,

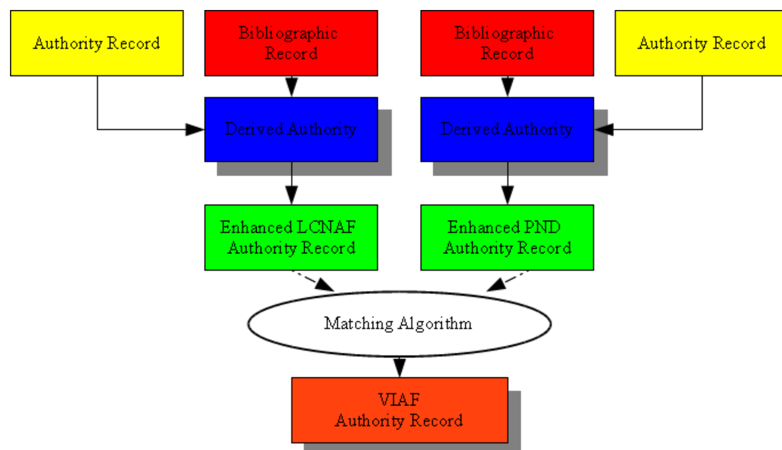


Figura 1: Algoritmo di schiacciamento in VIAF

per un punteggio tra zero e uno. Per calcolare la somiglianza di testo, veniva usato un trigramma basato su una tecnica di calcolo: i punteggi individuali erano modificati da un peso basato sulla "forza" (forte, moderato, debole) e poi sommati. Se il punteggio totale eccedeva una soglia predeterminata dal sistema durante la fase di testing, allora lo schiacciamento era confermato.

Tutte le tecniche descritte finora sono state usate per creare gli *enhanced autority record*, sia del PND che del LCNAF. La figura 1 mostra un semplice diagramma su come l'informazione fluisca prima di essere sottoposta al processo di schiacciamento.

Gli *enhanced autority record* di entrambe le fonti (LC e DNB) sono quindi il risultato del processo di revisione dei record provenienti dai propri authority file, sommati ai derived authority record, costituiti invece dalle informazioni supplementari estrapolate dal file bibliografico. Una volta costruiti entrambi *enhanced autority record*, questi vennero passati attraverso l'algoritmo di schiaccia-

mento: i record che ne derivarono, schiacciati e non, costituirono il primo nucleo del VIAF record. Nel 2006, nel VIAF file iniziale, erano presenti 6.300.000 enhanced authority record: fatta eccezione dei 500.000 record schiacciati, 3.700.000 record non schiacciati provenivano dal LCNAF, mentre 2.100.000 dal PND. Questi risultati si avvicinavano molto alle stime elaborate dagli studiosi basandosi sul test manuale; 250.000 coppie aggiuntive di authority record che rappresentano la stessa persona, ma che non potevano essere schiacciati automaticamente a causa della mancanza di informazioni addizionali utilizzabili (9). Per semplificare lo schiacciamento, tutto il testo dell'informazione viene normalizzato usando una versione modificata delle regole di normalizzazione NACO (Name Authority Cooperative Program of the Program for Cooperative Cataloging).[5] Il numero delle occorrenze di un termine particolare è archiviato nel sottocampo $9. Poiché questa informazione è predisposta al trattamento della macchina, essa non sarà necessariamente presente nella visualizzazione del record finale destinata all'utenza. Una volta formati, i VIAF authority record avrebbero incluso anche un VIAF record number sequenziale.

Dal punto di vista strutturale, poiché lo scopo iniziale del VIAF era di fornire un collegamento tra i file, il VIAF record non comprenderà mai al suo interno un campo 100 (Personal Name Heading), poiché non esiste un singolo nome autorizzato; ciascuna forma variante del nome sarà quindi registrata nel campo 700 (Heading Linking Entry), assieme a un'indicazione della sua fonte. Quando l'algoritmo determina uno schiacciamento, all'interno del record vengono predisposti due accessi per il collegamento. Quando un nome non viene schiacciato, nel record apparirà un unico campo 700. L'informazione supplementare è inclusa negli *enhanced authority record* come campo

[5] NACO, componente del Program for Cooperative Cataloging, è il programma ospitato dalla LC dedicato all'authority control di nomi, titoli uniformi e serie. http://www.loc.gov/catdir/pcc/naco.

locale (9xx). Quando nuove agenzie nazionali aderiranno al progetto VIAF, i loro authority file appena aggiunti saranno comparati con i record VIAF esistenti e migliorati, incorporando schiacciamenti aggiuntivi non appena questi vengono fatti all'interno dei record VIAF. Una volta ultimati, l'informazione migliorata proveniente dai record schiacciati viene fusa anch'essa. La versione del VIAF ha aiutato notevolmente i gestori del progetto, fornendo suggerimenti utili al fine di facilitare e migliorare la costruzione dell'authority file virtuale. Tra questi:

- se diminuissero i cambiamenti apportati agli authority record, lo schiacciamento migliorerebbe significativamente: molti schiacciamenti falliti erano infatti il risultato di un'analisi sbagliata della nota sulle fonti (campo 670);
- una struttura aggiuntiva che aiuti a evitare l'uso di nomi e titoli abbreviati sarebbe molto utile, come anche collegamenti espliciti al record bibliografico sorgente; l'inclusione della forma completa del nome dovrebbe comparire almeno come riferimento incrociato;
- utile sarebbe anche identificare esplicitamente il ruolo ricoperto dalla persona, e la sua specializzazione (compositore, illustratore, matematico etc.), al fine di migliorare ulteriormente lo schiacciamento, sia manuale che automatico.

Il VIAF oggi: utilizzo del servizio

Come ho già fatto rilevare precedentemente, il VIAF fornisce un servizio virtuale di ricerca di voci d'autorità, nella lingua, scrittura, alfabeto utilizzati dall'utente. Per questo motivo, chi potenzialmente può usufruire del servizio non è solo il bibliotecario, ma anche

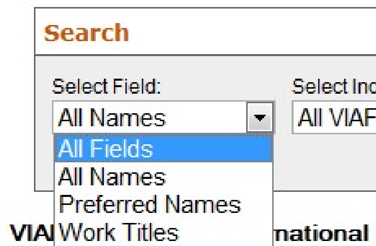


Figura 2: Dettaglio della schermata di ricerca

l'utente nel senso più generico. Per i bibliotecari che si occupano di authority control, il VIAF costituisce uno strumento assai utile in termini di risparmio di tempo nella ricerca della forma accettata di un nome: attraverso un'unica query, infatti, è possibile ottenere il record con la forma accettata del nome e tutte le sue varianti, evitando così lunghe ricerche in ogni singolo catalogo. Per l'utente generico, il VIAF potrebbe servire da semplice traduttore del nome desiderato nella propria lingua, scrittura e alfabeto, o in quelli degli altri cataloghi delle biblioteche partecipanti. Grazie all'utilizzo di UNICODE da parte del servizio, è possibile visualizzare il nome cercato in alfabeto latino, in cirillico, ebraico, arabo e in uno dei sistemi grafici giapponesi. Per accedere al VIAF è necessario collegarsi all'indirizzo http://www.viaf.org; il sistema funziona con la maggior parte dei browser Web. L'interfaccia con cui l'utente si trova a interagire è molto semplice, lineare e intuitiva; in testa alla homepage risiede il search box. La ricerca del nome desiderato può essere fatta secondo criteri diversi scelti dall'utente, che per il momento sono: - All fields: tutti i campi. Questo indice recupera i record che contengono il termine cercato in qualunque punto del VIAF record, inclusi i titoli, gli identificatori di record, gli ISBN e campi nome. - All names: tutti i nomi. La ricerca è ristretta ai campi

nome e include sia le forme accettate che alternative. - Preferred names (lett.: nomi preferiti), cioè le forme accettate; la ricerca viene effettuata solamente sulle forme accettate del nome. - Work titles: titoli dell'opera. La ricerca del nome avviene nel campo titolo. Utilizziamo un esempio per analizzare nel dettaglio le varie casistiche offerte da una ricerca effettuata nel VIAF. Scrivendo Mozart nel search box otteniamo 5923 risultati, ciascuno identificante un VIAF record differente. L'ordine d'inserimento delle parole non è importante: una ricerca per Mozart, Amadeus troverà gli stessi record di un'altra ricerca fatta per Amadeus Mozart, o semplicemente Mozart: l'utente non è tenuto a conoscere i metodi di ricerca prettamente biblioteconomici, effettuando quindi la ricerca e usando i termini e i modi che predilige.

Dal marzo 2010 è stata inserita un'altra utilità di ricerca, ovvero un menù a tendina a creazione automatica, non appena si inizia a digitare nel search box il nome desiderato; a cascata, vengono forniti suggerimenti di ricerca. Vengono poi visualizzati i primi 100 risultati, secondo un ordine decrescente che predilige il nome con più dati legati, come mostra la figura 4 nella pagina successiva. In questo modo abbiamo ottenuto dei nomi che possiedono la stringa Mozart in un qualche punto del VIAF record. Muovendo il mouse sopra le bandierine collocate dopo il nome, possiamo vedere quali istituzioni abbiano scelto quella forma particolare del nome. Le bandierine con un piccolo punto esclamativo su di esse rappresentano record in fase di test. Il punto d'accesso numero 1 dell'esempio mostra quattro forme accettate del nome; ciascuna è seguita da una o più bandiere rappresentanti la fonte dell'intestazione, unita a un titolo esemplare, se disponibile. Il record "Barth, Karl", apparentemente non inerente alla ricerca, viene incluso nei risultati poiché il teologo svizzero ha scritto un'opera su Mozart: per questo motivo la stringa Mozart è presente nel VIAF record di Barth. Come già detto, la ricerca di

Figura 3: Suggerimenti di ricerca

Figura 4: Risultati di ricerca

un nome è effettuabile in specifici campi del record. Nel caso della ricerca Mozart, troveremo:

- 5923 VIAF record in All fields;
- 73 V. r. in All names;
- 56 V .r. in Preferred Names;
- 5813 V. r. in Work titles.

Una volta selezionato un nome, l'utente accede al VIAF record; questo si configura come una sorta di riassunto dei metadati legati a ciascun authority record originale e ad altre informazioni che potrebbero risultare utili: titoli associati con l'entità, editori, date di pubblicazione e, ovviamente, i riferimenti incrociati che mostrano le forme alternative del nome. Utilizzando il nostro esempio su Mozart, entrati nel record otterremmo una schermata simile a quella mostrata in figura 5 a fronte. Nella schermata troviamo, quando disponibile, una widget,[6] che mostra una selezione di immagini delle coperte dei volumi utilizzati come fonti delle notizie di authority, con la possibilità di scorrerle. Sotto di esso, viene presentato nel dettaglio l'elenco delle forme accettate (in formato UNIMARC o MARC-21) di ciascuna fonte, con la possibilità di entrare e visualizzare l'authority record specifico. Accanto all'elenco è presente un diagramma che mostra le relazioni tra i record delle istituzioni partecipanti a quel record: muovendo il cursore del mouse sopra i nodi del diagramma (le bandierine) o sui nomi della lista, delle linee di connessione mostrano delle informazioni circa gli schiacciamenti. Il diagramma presentava dei problemi di funzionamento con browser

[6]Nell'ambito della programmazione, la widget è una componente (tipicamente grafica) di un'interfaccia utente di un programma, finalizzata a facilitare l'interazione con il programma stesso. In italiano è detto congegno (o elemento) grafico. Tipici esempi di widget sono la Barra dei menù, il Menù a discesa, la Scrollbar etc.

Figura 5: Visualizzazione di un record

come Internet Explorer, non ancora ben risolti: tutt'oggi, a distanza di quattro anni, persiste molta lentezza nella visualizzazione degli schiacciamenti e degli authority file partecipanti al record, che a volte si trovano duplicati nel diagramma. Sotto i tre macroelementi appena esposti, appaiono un certo numero di sezioni. La sezione riguardante le Alternate Name Forms è dedicata alla visualizzazione di tutte le forme alternative del nome contenuto nel record. Rispettando le varianti nazionali di ciascuna agenzia partecipante al VIAF, vengono visualizzate le forme del nome nelle varie lingue, scritture e alfabeti. La sezione riguardante i Selected Titles (titoli selezionati) consiste in una lista di titoli di opere, tra i più comuni, associati al nome cercato: nel nostro caso, avremo titoli di opere associate a Mozart. Il numero tra parentesi, posto dopo ciascun titolo, rappresenta il numero delle occorrenze del titolo in riferimento a tutti i file. Countries of Publication mostra una mappa costruita sui codici dei paesi estrapolati dai record; la visualizzazione può essere personalizzata a seconda che si voglia focalizzare solo sull'Europa, solo sull'Asia o sul mondo intero. Publication Statistics mostra invece una linea temporale riferita alla storia "vissuta" dalla pubblicazione (figura 6 nella pagina successiva). Selected Publishers offre un riscontro immediato su quali agenzie possiedano opere di Mozart pubblicate da un particolare editore; il numero delle occorrenze è posto tra parentesi. About è la sezione dedicata alla descrizione dell'autore cercato. Dalla sezione ricaviamo che Mozart è un autore di sesso maschile, di nazionalità austriaca, e che la lingua utilizzata nelle sue opere è il tedesco. Vengono inoltre forniti dei link a risorse esterne, nel nostro caso Wikipedia e WorldCat Identities, dove potremmo ottenere ulteriori informazioni biografiche sull'autore; questa opzione soddisfa pienamente i requisiti richiesti dal web semantico, risultando oltremodo utile e soddisfacente per l'utente. La sezione History of VIAF mostra quali authority record siano stati aggiunti

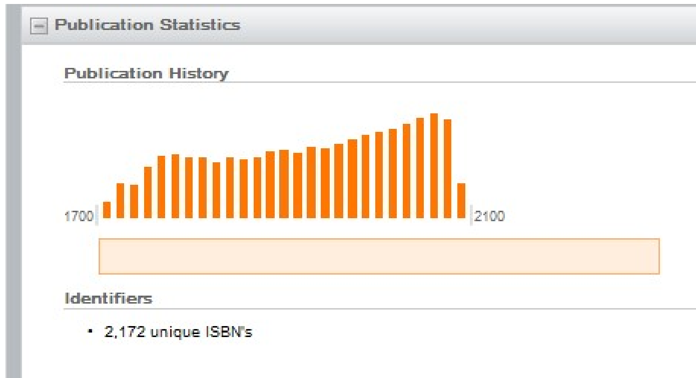


Figura 6: Statistiche



Figura 7: Informazioni sull'autore



Figura 8: Formati disponibili

al VIAF record, oppure cancellati in funzione di esso. La sezione Record views permette di accedere a diverse visualizzazioni del record nei formati attualmente disponibili (MARC-21, UNIMARC e RDF) grazie al sistema linked data:

Linked data è un sub-topic del Web semantico, ed è un termine usato per indicare un metodo di esposizione, condivisione e connessione di dati usando Uniform Resource Identifier (URI) e Resource Description Framework (RDF). Per i gestori del VIAF, i linked data vanno costruiti sulle prescrizioni dettate da Tim Berners-Lee (*Design Issues: Architectural and philosophical points*), e quindi:

- URIs per qualsiasi cosa possa essere identificata come risorsa;
- http 303 redirects per URIs che rappresentano le *personae* inerenti al nostro metadato;
- http content negotiation per diversi formati di dati;
- una visualizzazione RDF dei dati;
- un ricco set di collegamenti esterni e interni nel nostro record.

La visualizzazione completa del record in MARC-21, UNIMARC e RDF è fondamentalmente possibile grazie a *http content negotiation*. L'URI attuale restituisce una versione XML del record VIAF, il

quale viene restituito a sua volta nel browser in formato HTML via XSLT (eXtensible Stylesheet Language Transformations, linguaggio di trasformazione dell'XML). Non appena un maggior numero di authority file partecipanti al VIAF diverranno disponibili nel Web, i membri di OCLC introdurranno un maggior numero di link a siti esterni. Per quanto riguarda il formato RDF, questo viene restituito come un alquanto "rudimentale" Friend of a Friend (FOAF)[7] record. Poiché ci si aspetta che il VIAF supporti presto altre entità oltre alle persone (enti, famiglie, titoli, luoghi geografici), FOAF risulta abbastanza limitativo per questi scopi, aggiungendo il fatto che il match tra i concetti FOAF e i tipici authority file delle biblioteche non è molto buono (Hickey, «VIAF and RDF»). La LC stava aggiornando, verso la fine del 2009, le LCSH (Library of Congress Subject Headings) a SKOS (Simple Knowledge Organization System); i gestori del VIAF pianificarono un qualcosa di simile per il vocabolario dei soggetti FAST,[8] individuando come SKOS sia un'applicazione di RDF appetibile per essere usata dal VIAF. Sfortunatamente, SKOS non possiede la nozione di persona, e quindi manca di alcune interessanti classi RDF che invece FOAF possiede, come ad esempio

[7] FOAF è un'ontologia comprensibile dal computer volta a descrivere persone, con le loro attività e le relazioni con altre persone e oggetti. È un'estensione di RDF; le macchine possono usare FOAF, ad esempio, per cercare tutte le persone che vivono in Europa, o tutte le persone che hanno un amico in comune, proprio perché permette di definire le relazioni tra persone. Ogni profilo possiede un identificativo univoco (email, URI, homepage, blog, codice fiscale etc.) che viene utilizzato per definire queste relazioni. Per maggiori informazioni vedi sito web: http://www.foaf-project.org.

[8] Faceted Application of Subject Terminology deriva dalle LCSH, ma funziona con una sintassi semplificata.

L'obiettivo del progetto è di sviluppare uno schema per soggetti basato sulle LCSH, adatto ai metadati, e che sia facile da utilizzare, capire e mantenere. La fonte primaria di dati per il lavoro di ricerca è stata il database WorldCat di OCLC, contenente record bibliografici dotati di milioni di intestazioni diverse, geografiche e tematiche (Dean). Vedi anche il sito web di OCLC: http://www.oclc.org.

l'abilità di specificare le date di nascita e di morte, o di rendere disponibile un link alle pubblicazioni della persona. Per questo motivo, l'approccio iniziale tenuto da OCLC, e in particolare da Thomas Hickey, è stato quello di intrecciare una descrizione SKOS e FOAF di persone contenute nel VIAF, ma questa idea risultava sbagliata poiché RDF non riusciva a comprendere quando le persone sono di due differenti rdf:type. Andy Houghton (Consulting Software Engineer di OCLC) propose la creazione di risorse multiple real-world-object, che potevano essere descritte separatamente come FOAF e SKOS, e poi collegate, ma sembrava una soluzione troppo complicata(Hickey, «VIAF's new linked data»).

Recentemente, nel maggio 2010, tale approccio è stato rivisitato e migliorato. OCLC ha ritenuto opportuno conservare la descrizione SKOS (convenientemente rivista), ma ha voluto anche descrivere il concetto VIAF come FOAF, ed esporre la maggior parte dei dati in una forma che ricordi maggiormente il VIAF "nativo". Tutto ciò è stato realizzato coniando nuovi URI, poi descritti da RDF.

Conclusioni

Il Virtual International Authority File sta evolvendo molto velocemente, migliorando il proprio servizio grazie alla partecipazione di un numero sempre maggiore di biblioteche, dislocate ormai in tutti i continenti. Senza il recente coinvolgimento nel progetto della Bibliotheca Alexandrina, o dell'istituto di ricerca giapponese NII, non sarebbe stato possibile ottenere le forme varianti di nomi negli alfabeti delle loro rispettive nazioni.

Il punto di forza del VIAF consiste in una sempre maggiore adesione, da parte delle biblioteche, al servizio che esso fornisce: in altre parole, più le agenzie catalografiche sono varie nella lingua, scrittura o alfabeto, maggiore sarà il numero di forme varianti legate

a un determinato record, che finirà inevitabilmente con l'essere arricchito.

Ma c'è anche un rovescio della medaglia: infatti come già ho notato, il VIAF si costruisce semplicemente sugli authority file già formati di ciascuna agenzia catalografica; ora, se un determinato authority file presenta carenze sotto il profilo dell'authority control, in quanto contiene record non corretti o incompleti, questi sono destinati a non comparire mai nel VIAF record; così, semplici ricerche di autori nazionali non sarebbero soddisfatte.

Concludendo, è bene sottolineare che OCLC e VIAF stanno svolgendo una funzione fondamentale nella diffusione del lavoro effettuato sui dati a un network level, contribuendo entrambi allo sviluppo e alla diffusione dell'International Standard Name Identifier (ISNI),[9] nuovo identificatore standard di nomi. ISNI fornirà un mezzo per identificare unicamente i nomi dei profili pubblici (publicly facing) di autori, compositori e altri creatori di risorse, personaggi immaginari e storici e titolari di diritti, in particolar modo gli editori. Un identificatore di autorità servirà come link per le occorrenze di un'identità attraverso i database presenti nel web, e renderà più facile relazionare i nomi usati dagli editori a quelli usati nelle biblioteche.

L'ISNI è operativo dal 2010 sotto un consorzio gestionale guidato dal CISAC (International Confederation of Societies of Authors and Composers), che si occupa di diritti e scambi con organizzazioni e biblioteche.

Gli ISNIs verranno inizialmente assegnati abbinando i record offerti dai membri del consorzio; sono stati condotti con successo una serie di test usando il file VIAF e i suoi processi di schiacciamento fondamentali.

[9] http://www.isni.org.

Riferimenti bibliografici

Bennett, Rick, et al. «Viaf (Virtual International Authority File): Linking the Deutsche Nationalbibliothek and Library of Congress Name Authority Files». *International Cataloguing and Bibliographic Control : Quarterly Bulletin of the Ifla Ubcim Programme* 36.1. (2007): 12. A stampa.

Berners-Lee, Tim. *Design Issues: Architectural and philosophical points*. 2009. http://www.w3.org/DesignIssues/LinkedData.html. Online.

Berners-Lee, Tim, James Hendler e Ora Lassila. «The semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities». *Scientific American*. (May 2001). http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C7084A9809EC588EF21. Online.

Dean, Rebecca J. «Il progetto FAST». *Authority Control. Definizione ed esperienze Internazionali: Atti Del Convegno Internazionale, Firenze, 10-12 Febbraio 2003*. A cura di Mauro Guerrini, Barbara B. Tillett e Lucia Sardo. Firenze: Firenze University Press, 2003. A stampa.

Guerrini, Mauro e Lucia Sardo. *Authority Control*. Roma: Associazione italiana biblioteche, 2003. A stampa.

Hickey, Thomas B. «VIAF and RDF». *Outgoing: Library metadata tecnhiques and trends*. (2010). http://outgoing.typepad.com/outgoing/2009/12/viaf-and-rdf.html. Online.

——. «VIAF's new linked data». *Outgoing: Library metadata tecnhiques and trends*. (2010). http://outgoing.typepad.com/outgoing/2010/05/viafs-new-linked-data.html. Online.

Tillett, Barbara B. «A Virtual International Authority File». Boston: IFLA, 2001. http://archive.ifla.org/IV/ifla67. Online.

——. «Access control: a model for descriptive, holding and control records». *Convergence: Proceedings of the Second National Conference of the Library and Information Technology Association, October 2-6, 1988, Boston*. Chicago: American Library Association, 1990. 48–56. A stampa.

——. «Authority Control on the Web». 2000. http://www.loc.gov/catdir/bibcontrol/tillett_paper.html, online.

Tillett, Barbara B, Renate Gömpel e Susanne Oehlschläger. «Ifla Cataloguing Principles: Steps Towards an International Cataloguing Code : Report from the 1st Ifla Meeting of Experts on an International Cataloguing Code, Frankfurt, 2003». München: Saur, 2004. A stampa.

Tillett, Barbara B., Jaesun Lee e Ana L. Cristán. *IFLA Cataloguing Principles: Steps Towards an International Cataloguing Code, 4 : Report from the 4th IFLA Meeting of*

Experts on an International Cataloguing Code, Seoul, Korea, 2006. München: Saur, 2007. A stampa.

Informazioni

L'autore

Giulia Manzotti
NATO Defense College, Roma
Email: g.manzotti@ndc.nato.int

Il saggio

ACKNOWLEDGMENT:

Data di submission: 2010-09-27

Data di accettazione: 2010-11-28

Ultima verifica dei link: 2010-11-30

Data di pubblicazione: 2010-12-15

